Genome-wide association study of seed coat color in sesame (*Sesamum indicum* L.)

Hongxian Mei
Henan Academy of Agricultural Sciences
Chengqi Cui
Henan Academy of Agricultural Sciences
Yanyang Liu
Henan Academy of Agricultural Sciences
Yan Liu
Nanyang Academy of Agricultural Sciences
Xianghua Cui
Zhumadian Academy of Agricultural Sciences
Zhenwei Du
Henan Academy of Agricultural Sciences
Ke Wu
Henan Academy of Agricultural Sciences
Xiaolin Jiang
Henan Academy of Agricultural Sciences
Yongzhan Zheng (sesame168@163.com)
Henan Academy of Agricultural Sciences https://orcid.org/0000-0002-4889-8054
Haiyang Zhang
Henan Academy of Agricultural Sciences

Research article

Keywords:

Posted Date: February 3rd, 2020

DOI: https://doi.org/10.21203/rs.2.18296/v2

Abstract

Background: Sesame is an important and ancient oilseed crop. Sesame seed coat color is an extremely important agronomic trait, and is related to biochemical functions involved in protein and oil metabolism, and antioxidant content. Because of its complication, the genetic basis of sesame seed coat color remains poorly understood.

Results: Genome-wide association study (GWAS) using 42,781 single-nucleotide polymorphisms (SNPs) was performed with a diverse association-mapping panel comprising 366 sesame germplasm lines in 12 environments. In total, 224 significant SNPs ($P < 2.34\times10^{-7}$) explaining approximately 13.34% of the phenotypic variation on average were identified, and 35 significant SNPs were detected in more than 6 environments. Out of 224 significant SNPs, 22 were located in the confidence intervals of previous reported quantitative trait loci. A total of 92 candidate genes were identified in the vicinity of the 4 SNPs that were most significantly associated with sesame seed coat color.

Conclusions: The results in this paper will provide new insights into the genetic basis of sesame seed coat color, and should be useful for molecular breeding in sesame.

Background

Sesame (*Sesamum indicum* L., 2n = 2x = 26), which belongs to the *Sesamum* genus of the Pedaliaceae family, is one of the earliest domesticated crops [1]. It is mainly planted in tropical and subtropical regions in Asia, Africa, and South America [2]. Compared with the seeds of other main oil crops, e.g., rapeseed (*Brassica napus*), soybean (*Glycine max*), peanut (*Arachis hypogaea*) and olive (*Olea europaea*), sesame seeds not only have the highest oil content, but also are rich in proteins, vitamins, and specific antioxidants such as sesamin and sesamolin [3,4]. Because of its high oil quality and high nutritive value, sesame seed is regarded as ‘the queen of oil seeds’ and one of the best choices for health foods [5].

Seed coat color is one of the most important agronomic traits of sesame. It is related to biochemical functions involved in protein and oil metabolism, antioxidant content, and disease resistance [6]. The natural color of mature sesame seeds is diverse, varying from black to white through different intermediates such as gray, dark brown, brown, pale brown, yellow and dirty white [1]. In general, pale-colored sesame seeds contain more oil than dark-colored ones [6, 7]. Therefore, white sesame seeds are usually used to produce oil, and black sesame seeds are favored as food and medication in China. Significant attention has been paid to the inheritance of seed coat color in sesame. Some early classical genetic studies have suggested that sesame seed coat color is determined by two genes [8], while other reports have indicated that the genetic basis of sesame seed coat color is far more complex, which may involve multiple genes and their interactions [9,10,11]. In recent years, the genotyping load and cost has been significantly reduced by the next-generation sequencing technologies [12], several high-density genetic maps have been developed and a large number of quantitative trait loci (QTLs) for agronomically

important traits have been identified in sesame [13,14,15,16,17], including QTLs for seed coat color [6,15,18]. However, QTL mapping efforts using the segregated progeny of a bi-parental cross only enable the detection of a subset of loci/alleles within the crop, and offer limited resolution owing to the small number of informative recombination events between linked genetic loci [19]. As an alternative approach to traditional QTL analysis, the genome-wide association study also called association mapping or linkage disequilibrium (LD) mapping, taking advantage of both the wide phenotypic variation and the high number of historical recombination events in natural populations, has been used for dissecting complex traits in crop species [20,21], such as rice [22], maize [23], soybean [24], cotton [25], and rapeseed [26]. As an orphan or neglected crop, GWAS analysis in sesame is still limited. Wei et al. re-sequenced 705 diverse sesame germplasm accessions and performed a comprehensive GWAS on 56 agronomic traits for the first time [27]. Using a subset of 400 accessions from the above population, Dossa et al. performed a large-scale GWAS on five traits related to drought tolerance [28].

In this study, seed coat color of an association-mapping panel comprising 366 sesame germplasm accessions was measured in 12 environments, and 42,781 SNPs were developed by using specific-locus amplified fragment sequencing (SLAF-seq) [29]. By performing a large-scale GWAS on seed coat color, the stable significantly associated SNPs were explored. These SNPs will play an important role in understanding the genetic basis of seed coat color in sesame.

Results

Phenotypic variations of sesame seed coat color

In this study, 366 diverse sesame lines, which were assembled into an association-mapping panel [30], were used to evaluate the phenotypic variation of seed coat color, and three color space values (L-value, a-value, and b-value) with 2 replications were analyzed in 12 environments. Descriptive statistics for seed coat color are presented in Table 1. Although continuous and wide phenotypic variations were observed, the L-value, a-value, and b-value did not fit the normal distribution ($P < 0.01$; Fig. 1). The L-value exhibited a wide range of 10.53 to 63.40, with the coefficient of variation (CV) ranging from 14.08–22.94% among different environments. Similarly, the a-value ranged from 0.08 to 11.22, with CV ranging from 24.07–37.40%, and the b-value ranged from -0.47 to 18.75, with CV ranging from 15.51–24.50% (Table 1). Analysis of variance was performed for the seed coat color in 12 environments. The results showed that there were highly significant differences among genotypes (G), environments (E), and genotype × environment interaction (G × E) ($P < 0.01$). In addition, a correlation analysis showed significant correlations ($P < 0.01$) among different environments. The broad-sense heritability of the L-value was calculated to be 98.16%, while the broad-sense heritability of the a-value and b-value was 97.55% and 96.88%, respectively.

Genome-wide association analysis for sesame seed coat color

The analysis of genetic relatedness showed that the average relative kinship coefficient between any two inbred lines was 0.018 in the panel. Approximately 56.6% of the kinship coefficients were 0 (kinship

values = 0), 43.2% of the kinship coefficients ranged from 0 to 0.2 (0 < kinship values ≤ 0.2), and 0.2% of the kinship coefficients ranged from 0.2 to 0.5 (0.2 < kinship values ≤ 0.5) (Fig. 2). This pattern of genetic relatedness indicated that most lines in this study were distantly related, which may be attributable to the broad range of genotypes.

To uncover the genotypic variation of seed coat color in sesame, GWAS, which took into account the population structure and familial relatedness, was performed to identify the associated SNPs. The QQ plot displayed in Additional file 4 showed that the model could be used to identify significant SNPs. Using a uniform Bonferroni threshold for significance of $P < 2.34\times10^{-7}$, 224 significant SNPs (L-value 38, a-value 17, and b-value 169) were identified in 12 environments, with an average phenotype variation explained (PVE) of 13.34% (Fig. 3). Among these SNPs, 35 were detected in more than 6 environments, 24 were detected in more than 8 environments, and 14 were detected in more than 10 environments (Additional file 1).

Regarding the L-value, 38 significant SNPs were detected on 5 linkage groups (LGs), and the PVE of each SNP ranged from 8.75% to 21.90% (Fig. 3). Among these SNPs, 5 were detected on LG1, 25 on LG2, 1 on LG5, 1 on LG7, and 6 on LG8. The Manhattan plots showed that the most significant SNP S1_6648896 on LG1 was detected in all 12 environments, explaining 8.79–21.90% of the total phenotypic variation. On LG2, 8 multi-environment significant SNPs (S2_12167303, S2_12178804, S2_12178823, S2_12194998, S2_12232894, S2_12232938, S2_12447358, S2_12247409) were significantly associated with L-value in 7, 8, 8, 8, 7, 10, 8, and 9 environments with an averaged PVE of 10.75%, 11.01%, 11.01%, 14.75%, 14.87%, 13.95%, 12.45% and 12.24%, respectively (Additional file 1).

Regarding the a-value, 17 significant SNPs were identified on LG2, LG3 and LG7, explaining 8.26–25.46% of the total phenotypic variation (Fig. 3). Of all the significant SNPs, S7_6839839 was detected in all 12 environments and was the most significant SNP in 9 environments, which explained 17.40% of the total phenotypic variation on average (Additional file 1).

Regarding the b-value, 169 significant SNPs distributing on LG1, LG2, LG3, LG4, LG5, LG6, LG7, LG8, LG9, LG10, LG11 and LG13 were identified, and explained 8.68–31.35% of the total phenotypic variation (Fig. 3). The Manhattan plots showed that 3 peaks on LG1, LG2, and LG8 were repeatedly detected in more than 6 environments. Nine significant SNPs were detected on LG1. The SNP S1_6648896 with the lowest *P* value on LG1 was detected in 9 environments and explained 12.93% of the total phenotypic variation on average. Seventy significant SNPs were detected on LG2. S2_12168663 with PVE of 13.24% and S2_12337057 with PVE of 14.43% were both detected in 7 environments. S2_12336812 with PVE of 14.10% was detected in 8 environments. S2_12167303 with PVE of 14.53% and S2_12247358 with PVE of 16.32% were detected in 9 environments. S2_12026452 with PVE of 15.79%, S2_12178804 with PVE of 13.46%, S2_12178823 with PVE of 13.46% and S2_12194998 with PVE of 21.01% were detected in 10 environments. S2_12015779 with PVE of 17.76%, S2_12015820 with PVE of 17.63% and S2_12247409 with PVE of 18.36 were detected in 11 environments. S2_12232894 with PVE of 18.75% and S2_12232938 with PVE of 19.50% were detected in 12 environments. On LG8, 4 multi-environment

significant SNPs (S8_7910606, S8_8220220, S8_8311600, S8_8313501) were significantly associated with b-value in 7, 6, 6, and 7 environments with an averaged PVE of 17.04%, 15.32%, 10.43% and 18.08%, respectively (Additional file 1).

Candidate genes associated with sesame seed coat color

To predict the putative genes associated with sesame seed coat color, we focused on the most reliable and stable peaks on different LGs, including S1_6648896, S2_12232938, S7_6839839 and S8_8313501 (Fig. 3). The haplotype analysis showed that the SNPs S1_6648896, S2_12232938 and S7_6839839 were all in genomic regions that were in state of linkage equilibrium, while S8_8313501 was involved in a 213-kb LD block (Fig. 4). Within the LD block (S8_8313501), or 99 kb either side of the SNPs (S1_6648896, S2_12232938 and S7_6839839), a total of 21, 20, 31 and 20 genes were identified, respectively (Additional file 2). Of the 92 genes, 29 had no definite annotation concerning their biological functions, and 10 were annotated as putative or probable proteins. The remaining 53 genes had domains of known functions (Additional file 2). Gene ontology analysis indicated that the functions of most genes were binding, catalytic reactions, transferase activity, transcription regulator and transporter. For example, SIN_1016759 (41.05 kb from SNP S2_12232938 on LG2) and SIN_1023237 (88.65 kb from SNP S7_6839839 on LG7) encode polyphenol oxidase (PPO) and laccase-3, respectively, which have the functions of catalytic reactions [15]. SIN_1006022 encodes a cytochrome P450 94A2, which have the function of catalytic reaction [31]. SIN_1023226 and SIN_1024895 encode a WRKY transcription factor 67 protein and a transcription factor basic helix-loop-helix (bHLH130) protein, respectively, which are transcription regulators [32].

Discussion

GWAS has become an efficient and powerful tool at identifying genetic variations and loci responsible for the agronomically important traits [21]. In 2015, a GWAS of oil quality and agronomic traits with 705 sesame lines identified several causative genes, demonstrating the feasibility of GWAS in sesame [27]. In the present study, the panel of sesame accessions with wide geographic distribution, plentiful phenotype variation, sufficient genetic variation and weak population structure is advantageous for GWAS implementation [30]. However, the reliability of GWAS is usually disturbed by phenotypic variance associated with the environment. Multi-environment program is a practical way to correct for this error [25]. In this study, the trait experiments were performed at 4 sites, and these 4 sites belong to 3 climate classifications, temperate monsoon climate (PY and SQ), subtropical monsoon climate (NY) and tropical marine monsoon climate (SY). There are large differences in geographic position and climate among these sites. The influences of the environment were effectively eliminated in the phenotyping from multiple plots and years. Abundant phenotypic variation and stable heritability of seed coat color are suitable to reveal its genetic basis. By phenotyping in 12 environment (4 plots for 2 to 4 years) with 2 replications in the current study, we successfully identified 35 significantly stable associate signals in

more than 6 environments. These stable significant signals could provide useful information for exploiting target genes and analyzing the genetic basis of sesame seed coat color.

In this study, 224 significant SNPs with rigorous *P* values ($P < 2.34\times10^{-7}$) were identified in 12 environments. Meanwhile, 35, 24 and 14 of 224 significant SNPs were repeatedly detected in more than 6, 8 and 10 environments, respectively. Among 224 significant SNPs, we also found that 25 SNPs were simultaneously significantly associated with the L-value and b-value, and 7 SNPs were simultaneously significantly associated with L-value, a-value and b-value, indicating that these SNPs have pleiotropic effects on different color space values. To further confirm these significant SNPs associated with seed coat color in this paper, we compared our GWAS results with QTLs from previous linkage studies. Wang et al. identified 4 QTLs (*qSCa-4.1/qSCb-4.1/qSCl-4.1*, *qSCa-8.1/qSCb-8.1/qSCl-8.1*, *qSCl-8.2*, and *qSCb-11.1/qSCl-11.1*) for seed coat color in a RIL population [15]. Most of QTLs (3/4 QTLs) were verified by significant SNPs in the present study (Additional file 3). Eighteen significant SNPs on LG2 were mapped to the confidence interval of the QTL *qSCa-4.1/qSCb-4.1/qSCl-4.1*. One significant SNP (S1_6648896) and three significant SNPs (S1_9324398, S1_9330855 and S1_9332327) on LG1 were mapped to the confidence intervals of QTLs *qSCa-8.1/qSCb-8.1/qSCl-8.1* and *qSCl-8.2*, respectively. These comparison results corroborated our findings. Zhang et al. found 4 QTLs (*QTL1-1*, *QTL11-1*, *QTL11-2*, and *QTL13-1*) for sesame seed coat color [6], however, because of AFLP markers having been mainly used in the study of Zhang et al. in an independent genetic map, it is difficult to determine the relationship of the present loci to them. The remaining SNPs, which were not mapped to the confidence intervals of reported QTLs, indicated the likely existence of new seed coat color-related sites or environment-specific SNPs.

Researchers have made tremendous efforts regarding mapping major QTLs and identifying genes regulating seed coat color in diverse crop plants [33,34,35]. Seed coat color is determined by various plant secondary metabolites such as flavonoid compounds, including anthocyanin, flavonols, and proanthocyanidin, and possibly other phenolic relatives such as lignin and melanin [35]. PPOs, such as laccase, tyrosinase, and even peroxidase, have been reported to participate in the oxidation step in the biosynthesis of proanthocyanidin, lignin, and melanin, and produces black pigments via the browning reaction in plants [35,36]. In the present study, SIN_1016759 was located at 41.05 kb from S2_12232938, which was simultaneously significantly associated with the L-value (10 environments), a-value (2 environments), and b-value (12 environments). Homology analysis revealed that SIN_1016759 encodes a predicted PPO. Wei et al. reported that SIN_1016759 was strongly associated with seed coat color in sesame [27]. Wang et al. and Wei et al. showed that SIN_1016759 was located in the genomic region of a major QTL for seed coat color [15,18], and Wei et al. also indicated that SIN_1016759 was highly expressed in black sesame seeds from 11 to 20 days but not expressed in white sesame seeds [18]. Therefore, SIN_1016759 (*PPO*) might be the candidate gene controlling seed coat color in sesame.

Conclusion

In this study, GWAS for sesame seed coat color was performed using 42,781 SNPs with 366 sesame germplasm lines in 12 environments. A total of 224 significant SNPs were identified and 35 stable SNPs

were repeatly detected in more than 6 environments. Of 224 SNPs, 22 were located in the confidence intervals of reported QTLs, corroborating the GWAS results. Moreover, SNPs (S1_6648896, S2_12232938, S7_6839839 and S8_8313501) on 4 different LGs were the most reliable and significant loci, indicating that these loci contain important genes valuable for future research and breeding application. The GWAS showed great power in uncovering genetic variation in sesame seed coat color, and the results will provide new insights into the genetic basis of sesame seed coat color.

Methods

Plant materials and experiment design

In a previous study, 366 diverse sesame lines were selected from the Henan Sesame Research Center sesame germplasm collection, and were assembled into an association-mapping panel [32]. In this study, the panel was used for seed coat color evaluation and marker-trait association analysis.

The association-mapping panel was grown at four locations in China for two to four years: Nanyang (NY, E112.52°, N33.00°), from 2013 to 2014; Pingyu (PY, E114.63°, N32.97°), from 2013 to 2016; Shangqiu (SQ, E115.65°, N34.45°), from 2013 to 2014; and Sanya (SY, E109.50°, N18.25°), from 2012 to 2015. Field experiments were arranged by a randomized complete block design with 2 replication under every environment. Each accession was grown in a plot with 23–25 plants in a single row, with a distance of 0.15 m between plants within each row and 0.4 m between rows.

Measurement of seed coat color and statistical analysis

Sesame seeds were harvested from five randomly chosen plants in each line at maturity, and were mixed to evaluate the seed coat color. Seed coat color in each of 2 repetitions was scored using a HunterLab colorimeter (ColorFlex EZ, Hunter Associates Laboratory Inc., Virginia, USA), and was decomposed into L, a, and b color space values. The L-value represents brightness (black to white, 0 for black, 100 for white), the a-value represents the color from red to green (positive represents red, negative represents green), and the b-value represents the color from yellow to blue (positive represents yellow, negative represents blue) [37]. The seed coat color values were averaged over 3 technical repetitions. Descriptive statistics and Pearson correlation analysis, for sesame seed coat color value for each environment, were computed using the PROC UNIVARIATE and PROC CORR procedures (α = 0.01) of SAS8.02 software (SAS Institute, Cary, NC, USA), respectively. The analysis of variance was performed using QTL IciMapping V4.0 [38]. Broad sense heritability was calculated as: , where is the genotypic variance, is the genotype by environment variance, is the residual variance, k is the number of environments, and r is the number of replications [39].

Marker-trait association analysis

In a previous study, the association-mapping panel was genotyped by using SLAF-seq, and 89,924 high quality SNPs (minor allele frequency (MAF) ≥ 0.01 and integrity ≥ 0.7) were identified [32]. In this study, a

subset of 42,781 SNP markers with an MAF ≥ 0.05 and integrity ≥ 0.7 was used to calculate a principal components analysis (PCA) and perform marker-trait association analysis. PCA were calculated by using GCTA Software [40]. Marker-trait association analysis was performed using mixed linear models implemented in Tassel 5.0 software [41]. PCA and kinship were used to correct for false positives. The uniform Bonferroni threshold (negative log (0.01/n)) was used for the significance of associations between SNPs and traits, where n was the total number of SNP markers in the association analysis [42]. In this study, the threshold was 6.6 ($-\log_{10}(0.01/n) = -\log_{10}(0.01/42{,}781) \approx 2.34\times10^{-7} \approx 6.6$). Manhattan and QQ plots were drawn using the R package “qqman”.

Candidate gene prediction

To define the regions of interest for selection of potential candidate genes, the LD blocks, in which flanking SNP markers had strong LD ($r^2 > 0.6$), were defined as the candidate gene regions [43]. All genes within the same LD block ($r^2 > 0.6$) were considered as candidate genes. For significant SNPs outside of the LD blocks, the 99 kb (the LD decay distance) flanking regions on either side of the markers were used to identify candidate genes. LD heatmaps surrounding peaks in the GWAS were constructed using the R package “LDheatmap” [44].

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The data sets supporting the conclusions of this article are included within the article and its additional files.

Competing interests

The authors declare that they have no competing interests.

Funding

This study was supported by the China Agriculture Research System (CARS-14-1-01) and the Science-Technology Foundation for Outstanding Young Scientists of Henan Academy of Agricultural Sciences

(2018YQ27).

Authors' contributions

YZZ and HYZ conceived the study. HXM and YYL developed the association panel and supervised the experiments. CQC analyzed the data and drafted the manuscript, HXM revised the paper. YL, XHC, ZWD, KW and XLJ performed part of the library and field work. All authors read and approved the final manuscript. **Acknowledgements**

Not applicable

Abbreviations

CV, coefficient of variation

GWAS, genome-wide association study

LD, linkage disequilibrium

LG, linkage group

MAF, minor allele frequency

PCA, principal components analysis

PPO, polyphenol oxidase

PVE, phenotype variation explained

QTLs, quantitative trait loci

SLAF-seq, specific-locus amplified fragment sequencing

SNP, single-nucleotide polymorphism

References

1. Bedigian D, Harlan JR. Evidence for cultivation of sesame in the ancient world. Econ Bot. 1986;40:137–54.
2. Ashri A. Sesame breeding. In: Janick J, editors. Plant Breeding Reviews. New York:John Wiley & Sons Inc; 1998. p. 179–228.
3. Moazzami AA, Kamal-Eldin A. Sesame seed is a rich source of dietary lignans. J Am Oil Chem Soc. 2006;83:719–23.

4. Li C, Miao H, Wei L, Zhang T, Han X, Zhang H. Association mapping of seed oil and protein content in *Sesamum indicum* L. using SSR markers. PLoS ONE. 2014;9:e105757.
5. Wang L, Yu S, Tong C, Zhao Y, Liu Y, Song C, Zhang Y, Zhang X, Wang Y, Hua W, Li D, Li D, Li F, Yu J, Xu C, Han X, Huang S, Tai S, Wang J, Xu X, Li Y, Liu S, Varshney RK, Wang J, Zhang X. Genome sequencing of the high oil crop sesame provides insight into oil biosynthesis. Genome Biol. 2014;15:R39.
6. Zhang H, Miao H, Wei L, Li C, Zhao R, Wang C. Genetic analysis and QTL mapping of seed coat color in sesame (*Sesamum indicum* L.). PLoS ONE. 2013a;8:e63898.
7. Mei H, Wei A, Liu Y, Wang C, Du Z, Zheng Y. Variation and correlation analysis of sesamin, oil and protein contents in sesame germplasm resources. China Oils & Fats. 2013;38:87–90.
8. Brar GS, Ahuja KL. Sesame: its culture, genetics, breeding and biochemistry. In: Malik CP, editors. Annual Rev Plant Sci. New Delhi: Kalyani Publishers; 1979. p. 245–313.
9. Baydar H, Turgut I. Studies on genetics and breeding of sesame (*Sesamum indicum* L.) I. Inheritance of the characters determining the plant type. Turk J Biol. 2000;24:503–12.
10. Falusi O. Segregation of genes controlling seed colour in sesame (*Sesamum indicum* linn.) from Nigeria. Afr J Biotech. 2007;6:2780–3.
11. Pandey SK, Das A, Dasgupta T. Genetics of seed coat color in sesame (*Sesamum indicum* L.). Afr J Biotech. 2013;12:6061–7.
12. Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev Genet. 2011;12:499–510.
13. Zhang Y, Wang L, Xin H, Li D, Ma C, Ding X, Hong W, Zhang X. Construction of a high-density genetic map for sesame based on large scale marker development by specific length amplified fragment (SLAF) sequencing. BMC Plant Biol. 2013b;13:141.
14. Wu K, Liu H, Yang M, Tao Y, Ma H, Wu W, Zuo Y, Zhao Y. High-density genetic map construction and QTLs analysis of grain yield-related traits in sesame (*Sesamum indicum* L.) based on RAD-Seq techonology. BMC Plant Biol. 2014;14:274.
15. Wang L, Xia Q, Zhang Y, Zhu X, Zhu X, Li D, Ni X, Gao Y, Xiang H, Wei X, Yu J, Quan Z, Zhang X. Updated sesame genome assembly and fine mapping of plant height and seed coat color QTLs using a new high-density genetic map. BMC Genomics. 2016;17:31.
16. Zhang H, Miao H, Li C, Wei L, Duan Y, Ma Q, Kong J, Xu F, Chang S. Ultra-dense SNP genetic map construction and identification of *SiDt* gene controlling the determinate growth habit in *Sesamum indicum* L. Sci Rep. 2016;6:31556.
17. Mei H, Liu Y, Du Z, Wu K, Cui C, Jiang X, Zhang H, Zheng Y. High-density genetic map construction and gene mapping of basal branching habit and flowers per leaf axil in sesame. Front Plant Sci. 2017;8:636.
18. Wei X, Zhu X, Yu J, Wang L, Zhang Y, Li D, Zhou R, Zhang X. Identification of sesame genomic variations from genome comparison of landrace and variety. Front Plant Sci. 2016;7:1169.

19. Nordborg M, Weigel D. Next-generation genetics in plants. Nature. 2008;456:720–3.
20. Guo B, Wang D, Guo Z, Beavis W. Family-based association mapping in crop species. Theor Appl Genet. 2013;126:1419–30.
21. Huang X, Han B. Natural variations and genome-wide association studies in crop plants. Annu Rev Plant Biol. 2014;65:531–51.
22. Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, Li M, Fan D, Guo Y, Wang A, Wang L, Deng L, Li W, Lu Y, Weng Q, Liu K, Huang T, Zhou T, Jing Y, Li W, Lin Z, Buckler ES, Qian Q, Zhang Q, Li J, Han B. Genome-wide association studies of 14 agronomic traits in rice landraces. Nat Genet. 2010;42:961–7.
23. Li H, Peng Z, Yang X, Wang W, Fu J, Wang J, Han Y, Chai Y, Guo T, Yang N, Liu J, Warburton ML, Cheng Y, Hao X, Zhang P, Zhao J, Liu Y, Wang G, Li J, Yan J. Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. Nat Genet. 2013;45:43–50.
24. Zhou Z, Jiang Y, Wang Z, Gou Z, Lyu J, Li W, Yu Y, Shu L, Zhao Y, Ma Y, Fang C, Shen Y, Liu T, Li C, Li Q, Wu M, Wang M, Wu Y, Dong Y, Wan W, Wang X, Ding Z, Gao Y, Xiang H, Zhu B, Lee S, Wang W, Tian Z. Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. Nat Biotechnol. 2015;33:408–14.
25. Huang C, Nie X, Shen C, You C, Li W, Zhao W, Zhang X, Lin Z. Population structure and genetic basis of the agronomic traits of upland cotton in China revealed by a genome-wide association study using high-density SNPs. Plant Biotechnol J. 2017;15:1374–86.
26. Zhou Q, Han D, Mason AS, Zhou C, Zheng W, Li Y, Wu C, Fu D, Huang Y. Earliness traits in rapeseed (*Brassica napus*): SNP loci and candidate genes identified by genome-wide association analysis. DNA Res. 2018;25:229–44.
27. Wei X, Liu K, Zhang Y, Feng Q, Wang L, Zhao Y, Li D, Zhao Q, Zhu X, Zhu X, Li W, Fan D, Gao Y, Lu Y, Zhang X, Tang X, Zhou C, Zhu C, Liu L, Zhong R, Tian Q, Wen Z, Weng Q, Han B, Huang X, Zhang X. Genetic discovery for oil production and quality in sesame. Nat Commun. 2015;6:8609.
28. Dossa K, Li D, Zhou R, Yu J, Wang L, Zhang Y, You J, Liu A, Mmadi MA, Fonceka D, Diouf D, Cissé N, Wei X, Zhang X. The genetic basis of drought tolerance in the high oil crop *Sesamum indicum*. Plant Biotechnol J. 2019;17:1788–803.
29. Sun X, Liu D, Zhang X, Li W, Liu H, Hong W, Jiang C, Guan N, Ma C, Zeng H, Xu C, Song J, Huang L, Wang C, Shi J, Wang R, Zheng X, Lu C, Wang X, Zheng H. SLAF-seq: an efficient method of large-scale de novo SNP discovery and genotyping using high-throughput sequencing. PLoS ONE. 2013;8:e58700.
30. Cui C, Mei H, Liu Y, Zhang H, Zheng Y. Genetic diversity, population structure, and linkage disequilibrium of an association-mapping panel revealed by genome-wide SNP markers in sesame. Front Plant Sci. 2017;8:1189.
31. Le Bouquin R, Pinot F, Benveniste I, Salaün JP, Durst F. Cloning and functional characterization of *CYP94A2*, a medium chain fatty acid hydroxylase from Vicia sativa. Biochem Biophys Res Commun. 1999;261:156–62.

32. Li D, Liu P, Yu J, Wang L, Dossa K, Zhang Y, Zhou R, Wei X, Zhang X. Genome-wide analysis of *WRKY* gene family in the sesame genome and identification of the *WRKY* genes involved in responses to abiotic stresses. BMC Plant Biol. 2017;17:152.
33. Shao YF, Jin L, Zhang G, Lu Y, Shen Y, Bao JS. Association mapping of grain color, phenolic content, flavonoid content and antioxidant capacity in dehulled rice. Theor Appl Genet 2011;122:1005–1016.
34. Wang J, Xian X, Xu X, Qu C, Lu K, Li J, Liu L. Genome-wide association mapping of seed coat color in *Brassica napus*. J Agric Food Chem. 2017;65:5229–37.
35. Yu CY. Molecular mechanism of manipulating seed coat coloration in oilseed Brassica species. J Appl Genet. 2013;54:135–45.
36. Mayer AM. (2006). Polyphenol oxidases in plants and fungi: Going places? A review. Phytochemistry 67, 2318–2331.
37. Champa WAH, Gill MIS, Mahajan BVC, Aroraa NK. Postharvest treatment of polyamines maintains quality and extends shelf-life of table grapes (*Vitis vinifera* L.) cv. Flame Seedless. Postharvest Biol Tec. 2014;91:57–63.
38. Meng L, Li H, Zhang L, Wang J. QTL Icimapping: integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. Crop J 2015;3:269–83.
39. Kaler AS, Ray JD, Schapaugh WT, King CA, Purcell LC. Genome-wide association mapping of canopy wilting in diverse soybean genotypes. Theor Appl Genet. 2017;130:2203–17.
40. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011;88:76–82.
41. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics. 2007;23:2633–35.
42. Xu L, Hu K, Zhang Z, Guan C, Chen S, Hua W, Li J, Wen J, Yi B, Shen J, Ma C, Tu J, Fu T. Genome-wide association study reveals the genetic architecture of flowering time in rapeseed (*Brassica napus* L.). DNA Res. 2016;23:43–52.
43. Yano K, Yamamoto E, Aya K, Takeuchi H, Lo PC, Hu L, Yamasaki M, Yoshida S, Kitano H, Hirano K, Matsuoka M. Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. Nat Genet. 2016;48:927–34.
44. Shin JH, Blay S, McNeney B, Graham J. LDheatmap: an R function for graphical display of pairwise linkage disequilibria between single nucleotide polymorphisms. J Stat Softw. 2006; 16, Code Snippet 3.

Figures

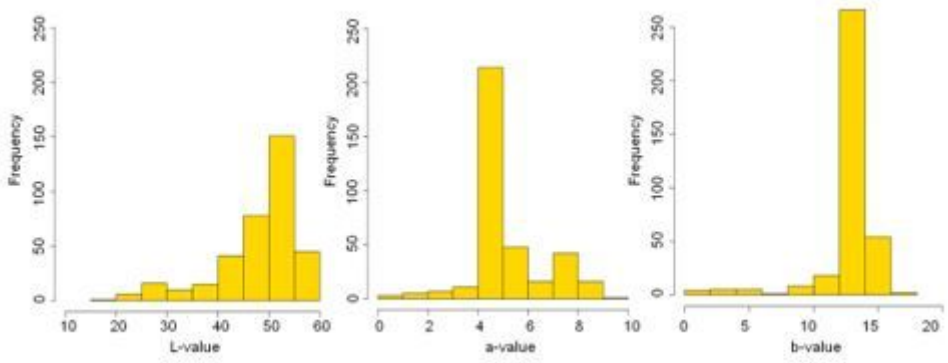

Figure 1

Frequency distribution of the average values of three traits related to sesame seed coat color

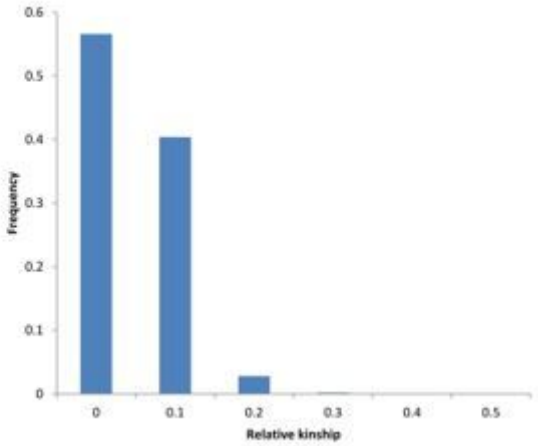

Figure 2

Analysis of relative kinships of the 366 sesame accessions

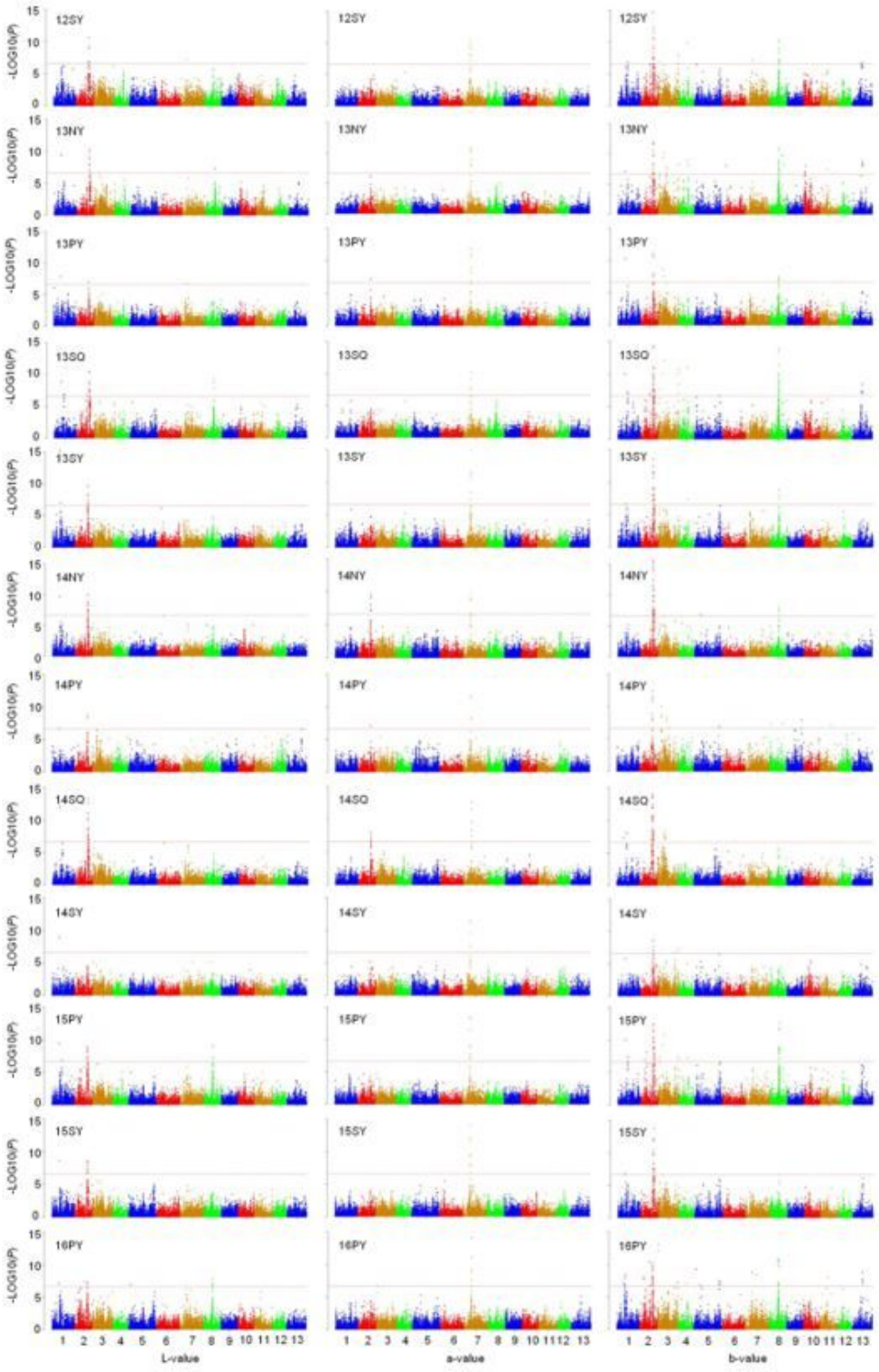

Figure 3

Genome-wide association studies (GWASs) of seed coat color in twelve environments

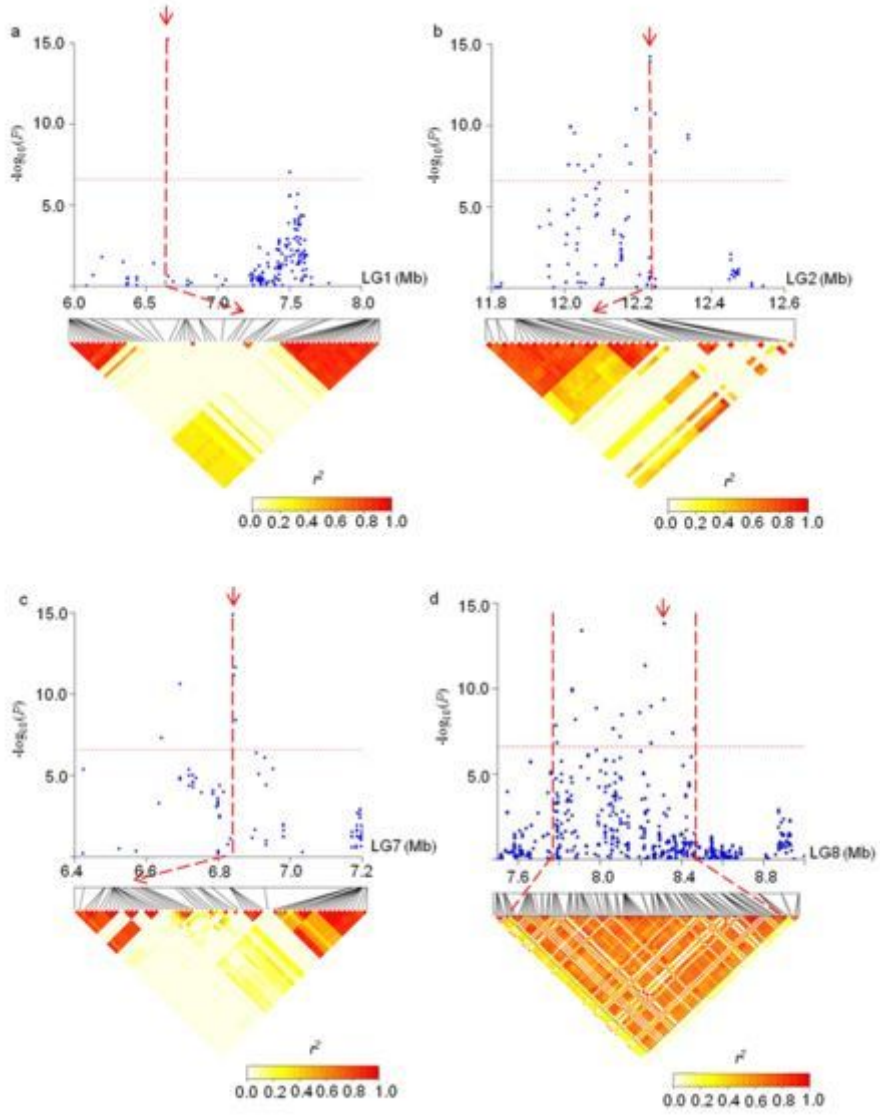

Figure 4

Local Manhattan plot (top) and LD heatmap (bottom) surrounding each peak on different linkage groups (LGs)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1TableS1.xlsx
- Additionalfile4FigS1.jpg
- Additionalfile2TableS2.xlsx
- Additionalfile3TableS3.xlsx